W474–W479 *Nucleic Acids Research, 2010, Vol. 38, Web Server issue* *Published online 18 May 2010*
doi:10.1093/nar/gkq407

MetaMHC: a meta approach to predict peptides binding to MHC molecules

Xihao Hu[1], Wenjian Zhou[1], Keiko Udaka[2], Hiroshi Mamitsuka[3,4] and Shanfeng Zhu[1,4,*]

[1]School of Computer Science and Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 200433, China, [2]Department of Immunology, Kochi Medical School, Nankoku, Kochi 783-8505, [3]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-0011 and [4]Institute for Bioinformatics Research and Development (BIRD), Japan Science and Technology Agency (JST), Japan

Received February 9, 2010; Revised April 13, 2010; Accepted May 2, 2010

ABSTRACT

As antigenic peptides binding to major histocompatibility complex (MHC) molecules is the prerequisite of cellular immune responses, an accurate computational predictor will be of great benefit to biologists and immunologists for understanding the underlying mechanism of immune recognition as well as facilitating the process of epitope mapping and vaccine design. Although various computational approaches have been developed, recent experimental results on benchmark data sets show that the development of improved predictors is needed, especially for MHC Class II peptide binding. To make the most of current methods and achieve a higher predictive performance, we developed a new web server, MetaMHC, to integrate the outputs of leading predictors by several popular ensemble strategies. MetaMHC consists of two components: MetaMHCI and MetaMHCII for MHC Class I peptide and MHC Class II peptide binding predictions, respectively. Experimental results by both cross-validation and using an independent data set show that the ensemble approaches outperform individual predictors, being statistically significant. MetaMHC is freely available at http://www.biokdd.fudan.edu.cn/Service/MetaMHC.html.

INTRODUCTION

The prediction of peptides that are presented by the restricting major histocompatibility complex (MHC) molecules is a crucial problem in immunology (1). MHC molecules bind short peptides derived from proteins in an allele-specific manner, and then present them on the surface of a cell for recognition by T-cell receptors (TCRs) (2). With the induction of the presented MHC-peptide complex, T cells proliferate and differentiate to help eliminate the antigens. As peptide presentation by MHC molecules is the prerequisite of cellular immune responses, it is of great importance to have the ability to accurately predict those peptides that bind to specific MHC molecules. This can help biologists and immunologists to elucidate the underlying mechanism of immune recognition as well as facilitating the process of epitope mapping and vaccine design (3). In contrast to biological experiments, computational approaches for predicting MHC binding peptides can significantly reduce the time and financial cost, which have been widely used to select a small number of candidate epitopes for experimental verification.

There are two major classes of MHC molecules, i.e. MHC Class I and MHC Class II molecules. MHC Class I molecules mainly present short endogenous peptides (around nine amino acids) to cytotoxic T cells (CTLs). In contrast, MHC Class II molecules mainly present longer peptides (usually 15–25 amino acids) from exogenous resources to helper T cells (Th). Since the binding groove of MHC Class II molecules is open at both ends, the location of the core binding motif in the peptide is highly variable, which makes predicting peptides binding to MHC Class II more challenging than predicting those binding to MHC Class I. Although a number of computational approaches have been proposed to address these problems, recent experimental results on benchmark data sets show that the improvement of predictive performance is needed, especially on the prediction of MHC Class II binding peptides (4–7). These computational approaches are usually based on different principles, such as position-specific scoring matrix (PSSM) (8–10), decision trees (11), artificial neural networks (ANN) (12), a stabilized matrix method (13,14), a virtual pocket matrix (15), hidden Markov models (16,17), support vector

*To whom correspondence should be addressed. Tel: +86 21 65643786; Fax: +86 21 65654253; Email: zhushanfeng@gmail.com machine (SVM) (18) and kernel-based methods (19), which may lead to quite different prediction results. On the other hand, because of the ability of integrating the performances of individual predictors, ensemble-based systems have been broadly deployed and achieved great success in a wide variety of areas (20).

The MetaMHC, which is also an ensemble-based web server for more accurate prediction of MHC-binding peptides, includes two components, MetaMHCI and MetaMHCII for the prediction of MHC Class I binding peptides and MHC Class II binding peptides, respectively. MetaMHC outperforms some popular prediction methods being statistically significant in both cross-validation and using an independent test data set.

METHODS

Workflow

The workflow of MetaMHC is shown in Figure 1. For each peptide sequence and a target MHC molecule, it first collects the prediction scores from several base predictors, which are then integrated as the final score by popular ensemble approaches.

Base predictors

To make the best use of ensemble approaches, the base predictors should be both accurate and diverse (20). Considering the recent performance evaluation results on benchmark data sets and the diversity of underlying prediction models (4–7), we choose ANN (21), SMM (21), NetMHC (22) and NetMHCPAN (23) as the base predictors in MetaMHCI, and SMM-align (14), TEPITOPE (15) and Local Alignment (LA) kernel (19) as the base predictors in MetaMHCII.

Integration strategies

MetaMHC implements four popular ensemble approaches for combining the results of different predictors. They are Consensus (6), PM (24), AvgTanh (25) and MetaSVMp, which is based on stacked generalization (26). The first two approaches have been already examined to achieve good performance in the prediction of MHC binding peptides (6,24), while the rest two have been found very successful in other applications of machine learning (25,26). The basic idea of each ensemble approach can be summarized as follows:

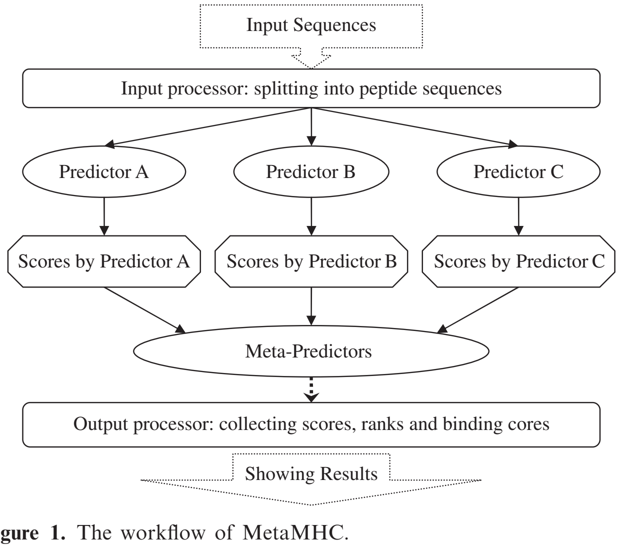


Figure 1. The workflow of MetaMHC.

- Consensus: a set of random peptides is collected as a reference list, and then each predictor ranks peptides in this reference list. For a test peptide, we can find one corresponding rank in the reference list by each predictor. The median rank by these methods will be given to this peptide as the consensus score. In MetaMHC, one million random peptides from the Swiss-Prot database are retrieved to generate a reference list.
- PM: the prediction scores of binders in training data by each predictor are assumed to obey a normal distribution. This is assumed for non-binders in training data as well, meaning that two distributions for binders and non-binders are generated by each predictor. For a test peptide, we first obtain the prediction score by each predictor, and then transform this score into the ratio of the probability of being a binder to the probability of being a non-binder by considering the two distributions of binders and non-binders. Finally, the product of all ratios over predictors will be given to the peptide as the final score.
- AvgTanh: the prediction scores of all peptides in the training data are assumed to obey a normal distribution. For a test peptide, after obtaining the prediction score by each predictor, we convert it into the *Z*-score, which is then normalized by the tanh function. The final score will be the average of all normalized scores by predictors. This strategy is not sensitive to outliers, because of the introduction of the tanh function.
- MetaSVMp: the prediction scores of all base predictor are used as the input of a support vector machine for predicting MHC binding peptides. One distinct advantage of this approach is that it can explore an nonlinear combination rule of the prediction results of base predictors.

Performance of MetaMHC

The performance of MetaMHCII on predicting MHC Class II binding peptides is evaluated on two recent benchmark data sets, (6) and (7), which we call the Wang data set and the Lin data set, respectively. Ensemble-based approaches achieve AUC (Area under ROC curve) of 0.72–0.83 for 11 human HLA DRB alleles in 10-fold cross-validation on the Wang data set, being significantly better than all base predictors: SMM-align, LA kernel and TEPITOPE (#peptides per allele: 245–3882; *P*-values <0.05 for all cases; Supplementary Material Table S1). We then apply the predictive model trained on the Wang data set to an independent test data set, the Lin data set and attained AUC of 0.69–0.89 for all six HLA DRB alleles, outperforming all base predictors except one case (#peptides per allele: 103; Supplementary Material Tables S2). Overall, MetaSVMp was the best ensemble strategy, being followed by AvgTanh. The performances of PM and Consensus were close to each other in the experiments. The improvement of MetaMHCII over individual predictors in terms of AUC on average ranges from 0.05 to 0.07 on the Wang data set, and from 0.01 to 0.08 on the Lin data set. Moreover, the performance of MetaMHCI has been validated in the recent Machine Learning in Immunology Competition (MLIC: http://www.kios.org.cy/ICANN09/MLI.html). At that time, MetMHCI consisted of three individual predictors, ANN, ARB and SMM (21). Both PM and AvgTanh achieved good prediction results in the competition, and outperformed two well-known predictors [BIMAS (8) and SYFPEITHI (9)] in all six categories (9- and 10-mer of three HLA molecules; Supplementary Material Table S3). Specifically, AvgTanh of MetaMHCI was awarded the winner in the category of HLA A*0101 of 9-mer and overall in the fourth place out of all 20 submissions in terms of average AUC in MLIC. All top three submissions (including one overall winner and four categories winners) were from Center for Biological Sequence Analysis (CBS) of Technical University of Denmark (DTU). Since the performance of MetaMHCI is directly correlated with the individual predictor of MetaMHCI, the worst performed task, B*0702 of 10-mer, may reflect the weakness of individual predictors. Due to the excellent performance of CBS, we thought that incorporating NetMHC and NetMHCPAN of CBS into MetaMHCI would further strengthen the predictive performance of MetaMHCI.

SERVER

Overview

MetaMHCI can make predictions on 82 different MHC I alleles, including 57 human alleles, 6 mouse alleles, 11 macaque alleles and 8 chimpanzee alleles (Supplementary Material Table S4). On the other hand, MetaMHCII can make predictions on 17 different MHC II alleles, including 14 human alleles and 3 mouse alleles (Supplementary Material Table S5). In addition to predicting MHC binding peptides using ensemble approaches, MetaMHC can also provide some useful links to other related prediction tools and databases/data sets. Moreover, help information on how to use MetaMHC is easily accessible on the web page.

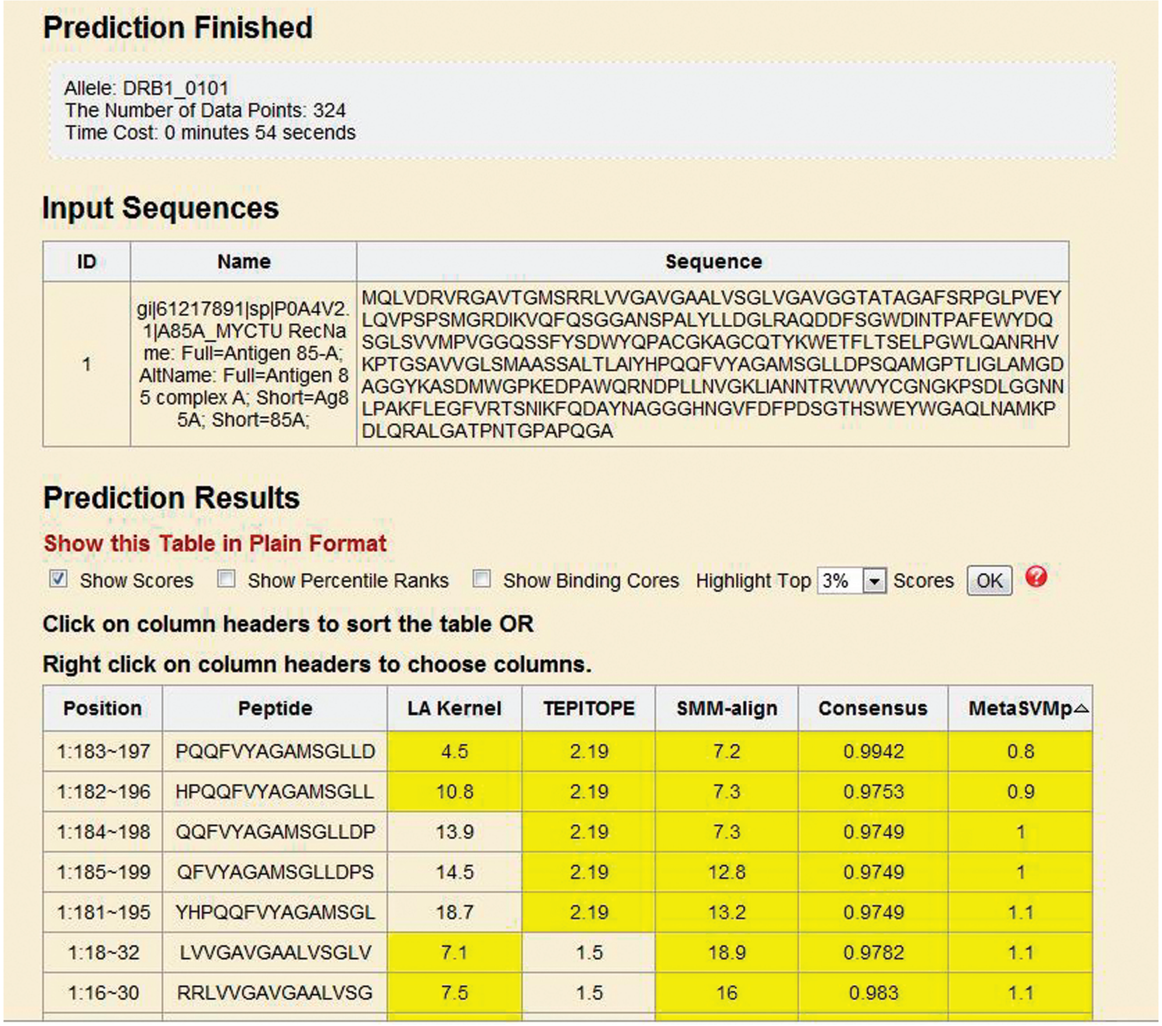
Input

MetaMHC predicts the binding affinity of all possible sub-peptides (according to the user-specified peptide length) hosted in one protein or multiple peptides. It accepts three types of input formats: the FASTA format of one protein, the plain format of one protein and the plain format of multiple peptides (with one peptide per line). The input data can be either pasted into the web interface directly or uploaded from a local file at the user's computer. The user also needs to specify some other information, such as the target MHC molecule, the base or ensemble predictors to be used, the peptide length for scanning the input protein, as well as the output format that can be either the web page or the plain text.

Output

The output interface of MetaMHC is illustrated in Figure 2. It first presents the name of the target MHC allele, the number of peptides predicted and the time spent during the prediction. The input sequence information is then displayed. If the FASTA format is used in the input, the corresponding name will appear in the result; otherwise a general name, such as 'Sequence 1' and 'Sequence 2', etc., is used instead. Finally, it comes to the main part of the output: the prediction result. The default output of the prediction result in the web interface is 'Show Scores', which shows the prediction score by each predictor. If 'Show Percentile Ranks' is chosen, the user can incorporate the top percentile ranks out of the prediction scores of one million random peptides from the Swiss-Prot database into the output. As illustrated in Figure 2, the prediction result is shown in a table format. The first column (Position) indicates the position where the peptide appears in the input sequence. For example, '1:183~197' means that the peptide starts at the 183th amino acid and ends at the 197th amino acid in the first protein. The second column (Peptide) is the primary sequence of the peptide. All other columns are the prediction results by selected prediction methods. In MetaMHCI, the prediction scores by three base predictors, ANN, SMM, NetMHC and NetMHCPAN, are all shown in values of IC50. On the other hand, in MetaMHCII, the prediction scores by LA kernel and SMM align are values of IC50, while that by TEPITOPE is not a value of IC50 but where a higher score means a better binding ability. The columns for ensemble predictors are on the right-hand side of the columns for base predictors. For Consensus, PM and AvgTanh, a larger score means stronger binding, whereas for MetaSVMp, a smaller score means stronger binding. We note that MetaSVMp can predict a value of IC50. From the result on performance evaluation of these ensemble approaches, AvgTanh in MetaMHCI and MetaSVMp in MetaMHCII are suggested to be the most favorable predictors. For MetaMHCII, the user can display the predicted binding core by choosing 'Show Binding Cores'. Additionally, the user can click 'Show this Table in Plain Format' to get the prediction result in the plain-text format for easy postprocessing.

In general IC50 of 500 nm is used to distinguish binders from non-binders. In addition, MetaMHC provides an easy way to highlight promising epitope candidates: if the predicted score of a peptide is lower than a threshold (3% for default), which corresponds to a percentile score of one million random peptides, the score will be highlighted for identifying the most probable epitope candidates easily. The user can change the threshold for a more strict or loose criterion. However, due to the limitation in the quality of training data and the accuracy of existing methods, the predicted percentile and the IC50



Prediction Finished

Allele: DRB1_0101
The Number of Data Points: 324
Time Cost: 0 minutes 54 seconds

Input Sequences

ID	Name	Sequence
1	gi|61217891|sp|P0A4V2.1|A85A_MYCTU RecName: Full=Antigen 85-A; AltName: Full=Antigen 85 complex A; Short=Ag85A; Short=85A;	MQLVDRVRGAVTGMSRRLVVGAVGAALVSGLVGAVGGTATAGAFSRPGLPVEY LQVPSPSMGRDIKVQFQSGGANSPALYLLDGLRAQDDFSGWDINTPAFEWYDQ SGLSVVMPVGGQSSFYSDWYQPACGKAGCQTYKWETFLTSELPGWLQANRHV KPTGSAVVGLSMAASSALTLAIYHPQQFVYAGAMSGLLDPSQAMGPTLIGLAMGD AGGYKASDMWGPKEDPAWQRNDPLLNVGKLIANNTRVWVYCGNGKPSDLGGNN LPAKFLEGFVRTSNIKFQDAYNAGGGHNGVFDFPDSGTHSWEYWGAQLNAMKP DLQRALGATPNTGPAPQGA

Prediction Results

Show this Table in Plain Format

☑ Show Scores ☐ Show Percentile Ranks ☐ Show Binding Cores Highlight Top 3% Scores OK

Click on column headers to sort the table OR

Right click on column headers to choose columns.

Position	Peptide	LA Kernel	TEPITOPE	SMM-align	Consensus	MetaSVMp△
1:183~197	PQQFVYAGAMSGLLD	4.5	2.19	7.2	0.9942	0.8
1:182~196	HPQQFVYAGAMSGLL	10.8	2.19	7.3	0.9753	0.9
1:184~198	QQFVYAGAMSGLLDP	13.9	2.19	7.3	0.9749	1
1:185~199	QFVYAGAMSGLLDPS	14.5	2.19	12.8	0.9749	1
1:181~195	YHPQQFVYAGAMSGL	18.7	2.19	13.2	0.9749	1.1
1:18~32	LVVGAVGAALVSGLV	7.1	1.5	18.9	0.9782	1.1
1:16~30	RRLVVGAVGAALVSG	7.5	1.5	16	0.983	1.1

Figure 2. A sample output of MetaMHC.

value may not be very precise, and thus need to be carefully explained. Another nice feature of MetaMHC is a customized output of prediction results. By default, results in the table are ranked by the starting positions of peptides. The user can sort the results in the table with any column in an ascending or descending order by clicking the header of the column of interest. Moreover, by right click on the header of the table, the user can hide or display any column of the table. These visualization techniques can help users explore the prediction results more conveniently.

IMPLEMENTATION

MetaMHC uses the JavaServer Pages (JSP) technology to handle the input validation, the output organization and the programming implementation of all predictors. JavaScript is included to provide interactive interfaces such as showing examples and creating dynamic tables. All web pages on MetaMHC are compatible with mainstream web browsers.

MetaMHC comprises the executable program provided by analysis tools of Immune Epitope Database (IEDB) to implement ANN and SMM and the executable program provided by CBS of Technical University of Denmark (DTU) to implement SMM align, NetMHC and NetMHCPAN (14,21–23). Regarding the SVM-based methods, i.e. LA kernel and MetaSVMp, MetaMHC uses the free package of LibSVM in the Java version (http://www.csie.ntu.edu.tw/~cjlin/libsvm/) to solve the optimization problems.

RELATED WEB SERVERS

A lot of web servers have been developed to address the problem of predicting MHC binding peptides. Lin *et al.* (5,7) have evaluated the performance of 27 servers for predicting peptides binding to MHC class I and 21 servers for predicting peptides binding to MHC Class II For predicting MHC Class I binding peptides, the best performed web servers include NetMHC (22), IEDB (SMM) and IEDB (ANN) (21). For predicting MHC class II binding peptides, the best performers include

NetMHCIIPAN (27), IEDB-consensus (21), PROPRED (15) and MULTIPRED (SVM) (28). MetaMHC make use of the most of these best performers as base predictors, such as IEDB (ANN), IEDB (SMM), NetMHC in MetaMHCI and TEPITOPE (using the same matrix as that of PROPRED) in MetaMHCII. Although NETMHCIIPAN is not included in the MetaMHCII, the performance comparison between MetaMHCII and NETMHCIIPAN on the Lin data set shows that each predictor outperforms the other one in three alleles (Supplementary Material Table S6). Furthermore, we emphasize that MetaMHC focuses on ensemble strategies to integrate various prediction results for better performance. Among four ensemble strategies in MetaMHC, Consensus is widely explored in the problem of predicting peptides binding to MHC, such as IEDB analysis tools (21). PM has been proposed by Karpenko *et al*. (24) to improve the MHC class II binding prediction, but there are no web servers that implement this strategy. In addition, MetaMHC is the first web server to implement AvgTanh and MetaSVMp for improving the performance of predicting MHC binding peptides.

CONCLUSIONS

Here, we present an ensemble-based web server, MetaMHC, for predicting MHC binding peptides. MetaMHC contains not only some popular base predictors, but also four types of ensemble strategies to improve the prediction accuracy significantly. With the accumulation of experimental data and the development of advanced prediction algorithms, such as NN-align (29), which is the latest method for predicting peptides binding to MHC Class II, MetaMHC will be regularly updated to explore a better performance for the problem of predicting peptides binding to MHC. A wide coverage of MHC Class I and II alleles and an easy-to-use web interface as well as the good prediction performance will make MetaMHC a very useful tool in the literature for peptide-based vaccine design.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Funding for open access charge: National Nature Science Foundation of China (No. 60903076 and 60773010); Shanghai Committee of Science and Technology, China (No. 08DZ2271800 and 09DZ2272800).

Conflict of interest statement. None declared.

REFERENCES

1. Lund,O., Nielsen,M., Lundegaard,C., Kesmir,C. and Brunak,S. (2005) *Immunological Bioinformatics*. The MIT Press, Cambridge, MA.
2. Janeway,C.A., Travers,P., Walport,M. and Shlomchik,M. (2001) *Immunobiology: The Immune System in Health and Disease*. Garland Publishing, New York.
3. Purcell,A.W., McCluskey,J. and Rossjohn,J. (2007) More than one reason to rethink the use of peptides in vaccine design. *Nat. Rev. Drug Discov.*, **6**, 404–414.
4. Peters,B., Bui,H.H., Frankild,S., Nielson,M., Lundegaard,C., Kostem,E., Basch,D., Lamberth,K., Harndahl,M., Fleri,W. *et al.* (2006) A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput. Biol.*, **2**, e65.
5. Lin,H.H., Ray,S., Tongchusak,S., Reinherz,E.L. and Brusic,V. (2008) Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol.*, **9**, 8.
6. Wang,P., Sidney,J., Dow,C., Mothé,B., Sette,A. and Peters,B. (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol.*, **4**, e1000048.
7. Lin,H.H., Zhang,G.L., Tongchusak,S., Reinherz,E.L. and Brusic,V. (2008) Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC Bioinformatics*, **9**, S22.
8. Parker,K.C., Bednarek,M.A. and Coligan,J.E. (1991) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *J. Immunol.*, **152**, 163–175.
9. Rammensee,H., Bachmann,J., Emmerich,N.P., Bachor,O.A. and Stevanović,S. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics.*, **50**, 213–219.
10. Bui,H.H., Sidney,J., Peters,B., Sathiamurthy,M., Sinichi,A., Purton,K.A., Mothé,B.R., Chisari,F.V., Watkins,D.I. and Sette,A. (2005) Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. *Immunogenetics*, **57**, 304–314.
11. Zhu,S., Udaka,K., Sidney,J., Sette,A., Aoki-Kinoshita,K.F. and Mamitsuka,H. (2006) Improving MHC binding peptide prediction by incorporating binding data of auxiliary MHC molecules. *Bioinformatics*, **22**, 1648–1655.
12. Nielsen,M., Lundegaard,C., Worning,P., Lauemoller,S.L., Lamberth,K., Buus,S., Brunak,S. and Lund,O. (2003) Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein. Sci.*, **12**, 1007–1017.
13. Peters,B. and Sette,A. (2005) Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC Bioinformatics*, **6**, 132.
14. Nielsen,M., Lundegaard,C. and Lund,O. (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*, **8**, 238.
15. Sturniolo,T., Bono,E., Ding,J., Raddrizzani,L., Tuereci,O., Sahin,U., Braxenthaler,M., Gallazzi,F., Protti,M.P., Sinigaglia,F. *et al.* (1999) Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nat. Biotechnol.*, **17**, 555–561.
16. Mamitsuka,H. (1998) Predicting peptides that bind to MHC molecules using supervised learning of hidden markov models. *Proteins*, **33**, 460–474.
17. Udaka,K., Mamitsuka,H., Nakaseko,Y. and Abe,N. (2002) Empirical evaluation of a dynamic experiment design method for prediction of MHC class I-binding peptides. *J. Immunol.*, **169**, 5744–5753.
18. Dönnes,P. and Kohlbacher,O. (2006) SVMHC: a server for prediction of MHC-binding peptides. *Nucleic Acids Res.*, **34**, W617–W622.
19. Salomon,J. and Flower,D.R. (2006) Predicting class ii mhc-peptide binding: a kernel based approach using similarity scores. *BMC Bioinformatics*, **7**, 501.
20. Polikar,R. (2006) Ensemble based systems in decision making. *IEEE Circuits Syst. Magazine*, **6**, 21–45.
21. Zhang,Q., Wang,P., Kim,Y., Haste-Andersen,P., Beaver,J., Bourne,P.E., Bui,H.H., Buus,S., Frankild,S., Greenbaum,J. *et al.* (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res.*, **36**, W513–W518.
22. Lundegaard,C., Lamberth,K., Harndahl,M., Buus,S., Lund,O. and Nielsen,M. (2008) NetMHC-3.0: accurate web accessible

predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res.*, **36**, W509–W512.

23. Nielsen,M., Lundegaard,C., Blicher,T., Lamberth,K., Harndahl,M., Justesen,S., Røder,G., Peters,B., Sette,A., Lund,O. *et al*. (2007) NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE*, **2**, e796.
24. Karpenko,O., Huang,L. and Dai,Y. (2008) A probabilistic meta-predictor for the mhc class II binding peptides. *Immunogenetics*, **60**, 25–36.
25. Jain,A., Nandakumar,K. and Ross,A. (2005) Score normalization in multimodal biometric systems. *Pattern Recogn.*, **38**, 2270–2285.
26. Wolpert,D.H. (1992) Stacked Generalization. *Neural Netw.*, **5**, 241–259.
27. Nielsen,M., Lundegaard,C., Blicher,T., Peters,B., Sette,A., Justesen,S., Buus,S. and Lund,O. (2008) Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol.*, **4**, e1000107.
28. Zhang,G.L., Khan,A.M., Srinivasan,K.N., August,J.T. and Brusic,V. (2005) MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Res.*, **33**, W172–W179.
29. Nielsen,M. and Lund,O. (2009) NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*, **10**, 296.